2020

Observing P300 Amplitudes in Multiple Sensory Channels using Cognitive Probing

Cody Lee Wintermute
Wright State University

OBSERVING P300 AMPLITUDES IN MULTIPLE SENSORY CHANNELS USING COGNITIVE PROBING

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science in Biomedical Engineering

by

CODY LEE WINTERMUTE

B.S.B.E., Wright State University, 2018

2020

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

29 July 2020

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY CODY LEE WINTERMUTE ENTITLED OBSERVING P300 AMPLITUDES IN MULTIPLE SENSORY CHANNELS USING COGNITIVE PROBING BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOMEDICAL ENGINEERING.

____
Sherif Elbasiouny, Ph.D.
Thesis Director

____
John C. Gallagher, Ph.D.
Interim Chair, Department of Biomedical Industrial, and Human Factors Engineering

Committee on Final Examination:

____
Sherif Elbasiouny, Ph.D.

____
Ulas Sunar, Ph.D.

____
Matthew Sherwood, Ph.D.

____
Barry Milligan, Ph.D.
Interim Dean of the Graduate School

ABSTRACT

Wintermute, Cody Lee. M.S.B.M.E., Department of Biomedical, Industrial, and Human Factors Engineering, Wright State University, 2020. OBSERVING P300 AMPLITUDES IN MULTIPLE SENSORY CHANNELS USING COGNITIVE PROBING

High cognitive workload occurs when excessive working memory resources have been deployed to resolve sensory and cognitive processing, resulting in decremented task performance. The P300 event-related potential (ERP) component has shown sensitivity to cognitive load, and it was hypothesized that an attenuated P300 amplitude could be indicative of high cognitive load. We tested this hypothesis by having eight participants complete two continual performance tasks at increasing workload levels while simultaneously performing an oddball task, evoking P300 ERPs in either the auditory or tactile sensory channel. In our experiment, electroencephalographic recordings were collected over the parietal region to observe the P300 component. Our results show a downward trend in P300 amplitude as workload increased when performing auditory oddball tasks, although P300's elicited by the tactile oddball tasks produced no consistent trend. These results suggest cognitive load indexing is possible in select sensory channels, though additional investigation is required.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

I. INTRODUCTION

1.1 Background

In an occupational setting, it is important for an employee to accomplish their goals in an efficient and timely manner. While any task will require work on the employee's part, excessive workloads can lead to elevated levels of stress which have been found to cause serious mental and physical health concerns (Teasdale, 2006). Unfortunately, this is not an uncommon situation for someone in the workforce. The American Psychological Association reported in 2018 that 42% of Americans identified heavy workloads as a significant contributor to their stress at work (APA, 2018). Further, this trend seems to be magnifying, as similar longitudinal surveying has shown a 20% increase in the number of workers claiming to be overworked over the last 30 years (Lipman, 2017). What practical measures can be taken to avoid stress, overwork, and injuries in the workplace? A possible solution may lie in monitoring the psychological barriers that define the limits of a person's capacity to do work.

1.1.1 Cognitive Load

Cognitive load is an abstract representation for an amount of working memory resources that have been allocated to complete a defined task. Any task that the brain attempts to complete – whether it be reading a book, merging onto the freeway, or picking out your outfit for the day – requires information to be stored and manipulated in working memory, which increments the amount of load being experienced (Miyake & Shah, 1999).

Cognitive Load Theory (CLT) was developed in the 1980's (Sweller, 1988), though prior research had alluded to a rudimentary understanding of the concept (G. A. Miller, 1956).

Within the CLT, the total cognitive load a task imposes on a person can be distributed into three categories: intrinsic, extraneous, and germane (Sweller, Van Merrienboer, & Paas, 1998). Any task has a base level of working memory associated with it, and this is the intrinsic cognitive load of the task. It is proportional to the complexity of the task. While any task will have some inherent difficulty associated with it, it can be further complicated by how the task is expressed. Instructions may not be clear, or the medium through which the task is expressed isn't optimal. The additional processing allocated to understanding how a task should be performed is referred to as extraneous load. Finally, the working memory that is reserved for organizing the learned information and connecting it to prior knowledge is known as the germane load.

Any action will have these three forms of cognitive load at varying levels. Understanding how these three types of load contribute to the overall load of a task is important when designing tasks that rely on differences in working memory and cognitive load. Intrinsic load is relatively static, but extraneous and germane loads can be increased or decreased depending on obstacles working against the participant or how the information and instructions are presented.

Cognitive load is a subjective measure of a person’s state and, other than the additive properties of the different types discussed, has no empirical method of being calculated. However, an arbitrary unit of measure can be ascribed to load in the form of working memory resources, which when “summed”, form a person’s working memory. This type of memory represents the information that can be recalled, retained, and applied simultaneously. For this reason, working memory can be classified as an executive function because no higher leveling reasoning can take place without the ability to hold and utilize information in working memory (Diamond, 2013).

A possible explanation for the mechanisms governing cognitive load and working memory is Multiple Resource Theory, defined in 1983 by Christopher Wickens. This theory posits that information processing from environmental stimuli can be performed more efficiently if the relevant stimuli are distributed among different sensory channels. In addition, the brain can allocate working memory resources in a similar manner. Under the MRT, these sensory channels do not compete against one another for working memory resources unless the demands of one channel are sufficiently high. In this situation, the brain can acquire resources it had deployed to one channel and divert them to channels that require additional resources. With deficient resources, neurological activity derived from these sensory channels will appear diminished (Wickens, Sandry, & Vidulich, 1983). This diversion of resources also results in decremented performance in stimulus-based tasks being performed (Basil, 2012).

Attention is another important factor in the processing of environmental stimuli. The distribution of attentional resources allows for parallel processing to occur when attending to multiple tasks. Capacity sharing allows for performance to remain optimal relative to the situation at hand (Kahneman, 1973; Tombu & Jolicœur, 2003). Similar to working memory resources, attentional resources are functionally finite for the optimal processing of sensory information, so the improper allocation of resources to task-irrelevant stimuli can lead to diminished performance as well (Broadbent, 2013; Pashler, 1984).

1.1.2 Assessing Cognitive Load

While it is difficult to measure directly, several different techniques have been devised to assess a person's cognitive load, and three primary domains have been used extensively: subjective techniques, performance-based techniques, and physiological techniques (Sweller et al., 1998). For example, the National Aeronautics and Space Administration developed its Task Load Index (TLX) assessment tool to subjectively evaluate a pilot or astronaut's cognitive load based on their perceived performance during a task (Hart & Staveland, 1988). The user rates themselves on six different subscales representing different aspects of cognitive load, as well as compares each subscale against the others to determine which were the greatest contributors. The Bedford Workload Scale was developed around the same time by the British Royal Aerospace Establishment and uses another subjective rating scale to determine cognitive load (Roscoe & Ellis, 1990). These subjective assessments are popular and effective for several reasons. They can be administered quickly, with almost no prep, instruction, or capital required. However, they have two main drawbacks: they are *post hoc* examinations, which prevents any online assessment, and their subjective nature can cause large between-subjects variability.

Other tests have been developed to combat the inherent biases in subjective assessment. Performance-based measures use the accuracy or efficiency with which a person completes a task as the basis for assessing cognitive load. When the total working memory capacity has been reached or exceeded during a task, performance will begin to degrade (S. Miller, 2001). If the primary task does not overstress the working memory system, a secondary task can be deployed in parallel to consume the unutilized working memory, which will then result in decreased performance in the primary task, the secondary task, or both (Wickens, Gordon, & Liu, 1998). Multiple Resource Theory predicts that this breakdown in performance will begin under less cognitive load if both tasks are competing for the same working memory resources, i.e. the tasks are targeting the same sensory channel (Derrick, 1988).

Finally, a third branch of assessment relies on the physiological responses of the human body to express cognitive load. The Index of Cognitive Activity analyzes disruptions in pupil size using eye-tracking equipment. It has been able to detect differences in cognitive load in real time while accounting for compounding effects, such as changes in lighting (Marshall, 2002). Other studies have shown success in using body temperature fluctuations and heart rate variability in discriminating between high and low task difficulty conditions (Haapalainen, Kim, Forlizzi, & Dey, 2010; McDuff, Gontarek, & Picard, 2014). Since these bodily reactions are largely independent of the person's thoughts and actions, they can be considered the least susceptible to subjectivity.

1.1.3 Using EEG to Assess Cognitive Load

Cognitive activity is an important physiological measure to the field of neuroscience, as it is a measurement of cortical potentials. The most pragmatic approach to this has been to record this brain activity using different imaging modalities. For example, functional near-infrared spectroscopy (fNIRS) – which measures variations of oxygenated hemoglobin in the cortical tissue to denote activity – has been shown to be sensitive to changes in cognitive load (Fishburn, Norr, Medvedev, & Vaidya, 2014). Functional magnetic resonance imaging (fMRI) has also been able to differentiate between levels of cognitive load (Callicott et al., 1999).

Electroencephalography (EEG) is another of these imaging modalities. EEG records electrical potentials present on the superficial layer of the cortex using scalp electrodes. Pyramidal neurons aligned orthogonal to the inner surface of the cortex generate post-synaptic potentials. These PSPs can propagate through the interstitial fluid of the scalp instantaneously, which lends EEG to precise temporal resolution and measurement. Individually, these PSP's have too low of a potential difference to register on an EEG recording, but the uniform orientation of the pyramidal neurons allows for the electrical dipoles to be summed, and differences can be detected when millions of these pyramidal neurons are firing simultaneously (Woodman, 2010).

Certain brain activity can be classified as rhythmic oscillations that are separated into distinct bands based on their frequency. EEG has been used to analyze frequency spectral patterns in brain activity to detect differences in cognitive load. Spectral power analysis of the different frequency bands has supported correlates in the 8-12 Hz alpha band (Anderson et al., 2011; Stipacek, Grabner, Neuper, Fink, & Neubauer, 2003), 4-8 Hz theta band (Anderson et al., 2011; Puma, Matton, Paubel, Raufaste, & El-Yagoubi, 2018), and the 1-4 Hz delta band (Harmony, 2013; Zarjam, Epps, & Lovell, 2015).

1.1.4 Event-Related Potentials

Frequency spectra are one way to extract meaningful data from EEG. Another is the event-related potential (ERP). These brain signals are generated as a response to explicit events, either in the environment or as the result of a cognitive process (Luck, 2005). The ERP was first observed in the 1930's (H. Davis, Davis, Loomis, Harvey, & Hobart, 1939; P. A. Davis, 1939) using a primitive form of EEG, but has expanded since then into an extensive subfield of neuroscience and cognitive psychology.

Within the EEG recording, there are countless neurological actions present, and isolating any single signal within the continuous data is difficult, though single trial ERPs are possible (Blankertz, Lemm, Treder, Haufe, & Müller, 2011; Jung et al., 1999). What is more common is to time-lock a selection of data epoched around an event, such as the presentation of an environmental stimuli, and then average across multiple trials. As more and more trials are averaged, the background noise will eventually zero out, and any ubiquitous signals associated with the chosen event will remain visible (Luck, 2005).

1.1.5 The P300

In particular, the P300b ERP component (P300) may be beneficial in monitoring cognitive load. The P300's name is derived from its nature as a positive-going voltage potential ("P") that classically peaks at 300 milliseconds ("300") post-stimulus onset, though this tends to range between 300-500ms (figure 1). The component is associated with reasoning and recognition of task-categorized target stimuli. In a scalp distribution map, the P300's largest amplitude is seen across the parietal region of the brain (Polich, 2007). The P300 has also been shown to be sensitive to cognitive load. As discussed earlier, changes in cognitive load manifest in different physiological responses to stimuli, and this can be observed in the morphology of event-related potentials. This has been observed with the P300 component. As workload increases, the amplitude of the component decreases (Donchin, 1986).

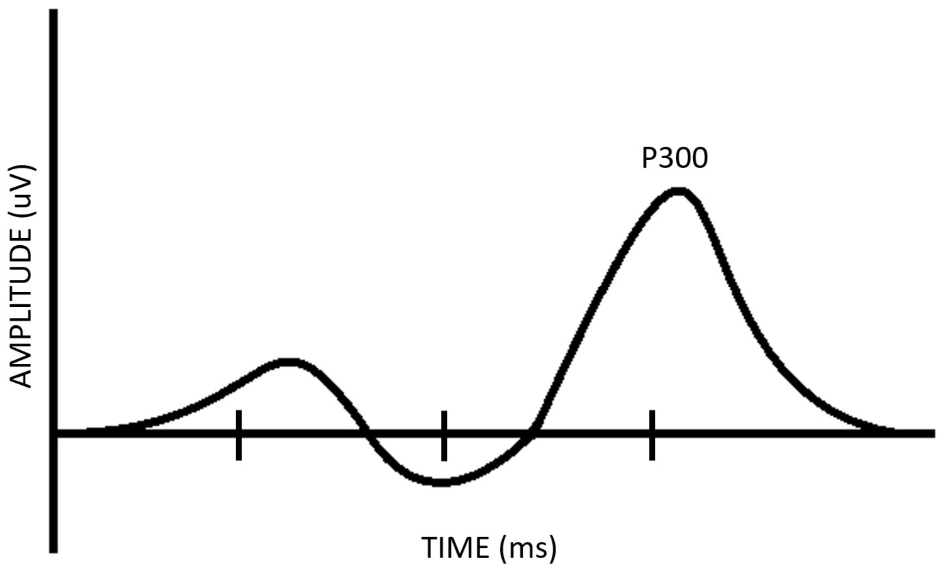

Figure 1: A stimulus-locked waveform of the P300b ERP component (Patel, Azzam, 2005)

One of the most common ways to evoke the P300 component is through the use of an oddball task. The oddball paradigm presents two unique stimuli to a participant in a pseudo-random order, with one rare oddball stimulus being presented fewer times than the more common standard stimulus (figure 2). When the oddball is categorized within the task as the target, the P300 is generated upon recognition of the oddball stimulus being presented. The morphology of the P300 is somewhat variable within the oddball paradigm, with both peak amplitude and peak latency able to be affected. P300 amplitude can be modulated by adjusting the rarity of the oddball stimulus. The less frequent the oddball is presented, the greater the amplitude of the component. Interestingly, it is the categories of stimuli (target, non-target) that matter more so than the individual distinct stimuli themselves (Kutas, McCarthy, & Donchin, 1977). It is important to note that there is evidence to suggest that the rarity of the target stimuli is a misnomer, and what is actually modulating component amplitude is the target-to-target interval, i.e. the time displacement between target stimuli (Croft, Gonsalvez, Gabriel, & Barry, 2003). In regards to the latency, it can be adjusted by altering the ease with which the two stimuli can be discriminated. In other words, the easier it is for the person to determine which stimulus is the oddball and which is the standard, the sooner the P300 component will peak in the time domain_(Magliero,_1984).

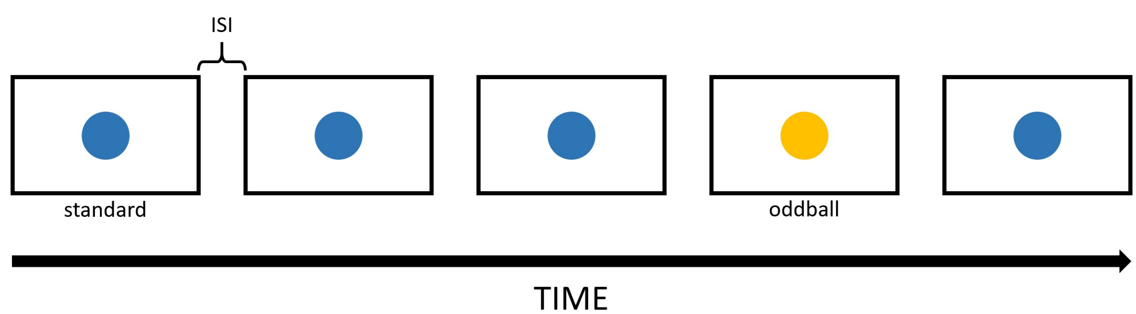

Figure 2: a schematic rendering of a visual oddball, showing the standard (blue) and oddball (yellow) stimuli

While it is common to see the oddball task deployed to target the visual sensory channel, other versions do exist that use different stimuli to generate the P300. For example, auditory oddball tasks can utilize tones at different frequencies in place of visual cues (Segalowitz & Barnes, 1993; Squires, Squires, & Hillyard, 1975), and tactile oddballs can use varying somatosensory stimuli to produce similar effects (Brouwer & Van Erp, 2010; Brouwer, van Erp, Aloise, & Cincotti, 2010; Herweg & Kübler, 2016). Even though the visual, auditory, and tactile stimuli are being processed in the occipital, temporal, and somatosensory regions, respectively, of the brain, all of the higher-level recognition processing is always most visible across the parietal region.

However, this should not imply that the area of the brain responsible for generating the P300 component is located in the parietal lobe. This is merely a result of the orientation of its electrical dipole, the propagation of the post-synaptic potentials, and the folding of the brain matter. The exact generator location for the P300 component (and many other ERP components) is unknown, though research suggests that the frontal lobe (Knight, 1984; Knight, Grabowecky, & Scabini, 1995) and the medial temporal lobe (Halgren et al., 1980; McCarthy, Wood, Williamson, & Spencer, 1989) are likely candidates. Further, having different sensory channels producing comparable brain signals compliments the assumptions made using the Multiple Resource Theory. In a dual task experimental design, an oddball task being deployed alongside another task will not compete for resources so long as the two tasks don't share a sensory channel, and one task does not become too difficult.

1.1.6 Cognitive Probing

Recently, new research utilizing passive brain-computer interfaces has suggested a novel method for interpreting cognitive load. Passive brain-computer interfaces (pBCI) provide support to a human-machine system by sending implicit inputs from neurological signals of the human to his or her machine teammate without generating an explicit output the user intends or even perceives. These neurological signals can be in reaction to cognitive probes the pBCI deploys in order to establish the user's cognitive state (Laurens R Krol & Zander, 2018). It can be beneficial to adapt a user's experience in a system with a pBCI, such as preventing task-irrelevant stimuli from distracting the user during high workload (Laurens Ruben Krol & Zander, 2017), thus conserving cognitive resources needed for the primary task. Cognitive probing is therefore a practical application of a phenomenon described in the Multiple Resource Theory.

1.2 Purpose

1.2.1 Risks Associated with High Cognitive Load

The concept of "choking" under pressure is a real phenomenon that is supported by the way in which high workloads can impact cognitive abilities. When modeling performance as a function of cognitive load, we see a nonlinear relationship with three distinct regions: underload, ideal load, and overload (figure 3). Performance is at its zenith when a person is aroused, alert, and attentive to the task at hand. The task is not necessarily simple or easy; in fact, it could be quite challenging, but the person can demonstrate competency and accuracy when incorporating an optimal proportion of working memory.

As the number of deployed working memory resources increases however, performance begins to wane. The person will become overloaded, a state where the person's working memory can no longer accommodate the task at hand. The person will become stressed and anxious, and the resulting drop in performance may exacerbate the situation, potentially causing a positive feedback loop resulting in total meltdown. Interestingly, poor performance is also associated with low cognitive load, a state known as underloading. This is when the person is not sufficiently engaged with the primary task. Boredom, apathy, and inattentiveness are displayed, and performance is decremented.

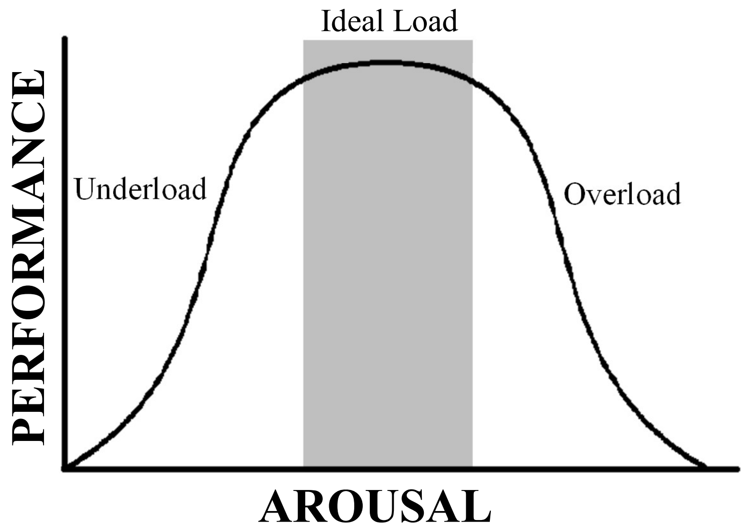

Figure 3: Stress response curve adapted from Yerkes-Dodson Model (Yerkes, Dodson, 1908)

Prior literature has shown that people who exhibit fewer working memory resources are more inclined to act impulsively and ascribe unnecessarily significance to certain information during decision-making (Burks, Carpenter, Goette, & Rustichini, 2009; Frederick, 2005). The consequence of this relationship is that unproductive or irrational behavior is correlated with a state of high cognitive workload. It is then important to state how this type of behavior may be counterproductive or even dangerous in a work-related context. High-stress professions, such as heavy-machinery operators, pilots, and surgeons, expose their partakers to excessive mental and physical demands that could potentially overload them and result in serious adverse consequences if performance dropped too low (Lindblom & Thorvald, 2014). Symptoms of this overloading are disruption of working memory, tunnel vision, and spatial unsteadiness (Sandblad, Lind, & Nygren, 1991). It is thus important to recognize when a person is becoming overloaded in the course of their duties, as well as correct this state of high cognitive workload.

The goal of this study is to observe changes in P300 morphology while participants complete a dual task cognitive probing experiment. The primary task will be a continuous performance task with specified low and high task load conditions. This primary task will target the visual sensory channel. The secondary task will be an oddball paradigm which targets either the auditory or tactile channels. It operates in the same capacity as a cognitive probe does in a passive brain computer interface. A control task where only the secondary oddball task is performed will also be deployed.

The intrinsic workload present in the secondary task is assumed to be constant across the control, low, and high task load conditions, and so it can be excluded from the total cognitive load being experienced during each trial. The primary task defines the task load level, and when workload becomes sufficiently high, the brain can no longer accommodate channels that are secondary to the primary tasks, thus causing "competition" between the multiple senses for resources. Resources will be acquisitioned to bolster sensory channels that are being overloaded, thus causing a decrease in performance output in the resource-starved channels. Given this relationship between mental workload, resource management, and cognitive probing, it can be conjectured that a drop in P300 amplitude during multi-task conditions can be quantified if the secondary channels are being targeted by a cognitive probing task.

1.3 Hypothesis

We hypothesize that the oddball P300 could be used to index the relative cognitive workload a participant is undergoing using this cognitive probing technique and that this index can be generalized to multiple different sensory channels. Primary and secondary task performance and subjective evaluations will also be used to compare against the ERP data.

A decrease in P300 peak amplitude is predicted to be observed between low and high workload conditions. No significant difference is expected in P300 peak latency. Primary and secondary task performance is predicted to decrease between the two load conditions in their respective metrics. Perceived cognitive workload in participants reported in subjective evaluation is expected to increase. No differences between targeted sensory channels were hypothesized.

II. METHODS

2.1 Materials

2.1.1 Participants

Each participant attended two separate sessions of data collection. On the first day, referred to as *Training*, the participant consented to participate and was given the opportunity to practice the different task conditions. Each task had a minimum performance threshold that the participant had to achieve to qualify for the study. If the threshold was met for all of the task conditions, the participant was asked to return one week later at the same time. On this second day, referred to as *Testing*, participants were asked to perform each of the task conditions in a repeated block design. The block was repeated four times. The study was approved and overseen by Institutional Review Boards representing both Wright State University and the Air Force Research Laboratory.

Thirteen participants were recruited for this study. Four participants did not complete the entire experiment. Two did not meet the performance requirements during training, one did not return for the second session, and one voluntarily ended the second session pre-maturely. Of the nine complete data sets, one was excluded due to poor data quality. In total, eight subjects (four female) between 18 – 42 years of age (mean = 25.13 years, SD = 7.9 years) were recruited for the study. All participants were given a pre-screening to exclude anyone with sensory impairment or psychiatric disorders. During this screening, each subject provided informed consent to participate. All participants were paid $15 per hour.

2.1.2 Recordings

Physiological recordings were performed using the BioSemi ActiveTwo system (BioSemi B.V., Amsterdam, The Netherlands). Recordings were made with a 2048 Hz sampling rate at 64 channel locations based on the modified combinatorial nomenclature extension of the 10-10 system (American Electroencephalographic Society, 1994) excluding the inferior chain with the exception of P9/P10 and Iz (Seeck et al., 2017), with bilateral electrodes on the mastoid process, infraorbital, and outer canthus locations. Additionally, a respiration band and two GSR electrodes were used to record respiration patterns and galvanic skin response, respectively. Participant responses were recorded using a low-latency mechanical keyboard (Cherry MX 6.0 [G80-3930], Cherry GmbH, Auerbach in der Oberpfalz, Germany).

2.1.3 Stimuli and Equipment

All tasks were coded in MATLAB (R2011b; The MathWorks, Inc., Natick, MA, USA) using the Psychophysics Toolbox (v3.0.13) (Brainard, 1997 ; Pelli, 1997; Kleiner et al., 2007). Stimuli were presented on a 24.5", 240 Hz monitor (BenQ ZOWIE XL2540, BenQ Corporation, Taipei, Taiwan) while participants sat approximately 65 cm away from the screen. The testing environment was a dark room with a natural sound machine (Dohm Classic, Marpac LLC, Wilmington, NC, USA) to cover background sounds. During certain tasks, participants used a game controller (Logitech F310 GamePad, Logitech, Newark, CA) to move a cursor around the screen to follow a target. The left or right analog stick was used during the task to correspond with the dominant hand of the participant.

The tactile oddball task used haptic stimulation in the form of vibrating motors to create different stimuli. Creating and recording this stimulation, particularly in the context of time-locking the events, presented a challenge as the BioSemi ActiveTwo systems used in this experiment were configured to input, output, and record audio signals, but not tactile signals. Knowing this, it was determined that an audio signal could be "hijacked" before it reached the speakers, transformed, and then recorded by the BioSemi.

To accomplish this, a full bridge rectifier circuit was built to convert the AC audio signal into a DC square wave signal. This was then sent to an Arduino microcontroller (Arduino Uno, Arduino, Somerville, MA) that was programmed to output the DC signal to two pager motors that were attached to the back of the subject's non-dominant hand. The amplitude of this signal was determined by the volume of the signal defined in the task code, which in turn controlled the amount of voltage being pushed to the motors. As such, the intensity of the vibrations felt by the participant was governed by the volume of the audio signal originally generated by the task code. Finally, the new tactile signal was returned to the input channel of the BioSemi, allowing for stimulus event markers to be generated in the recording software (figure 4).

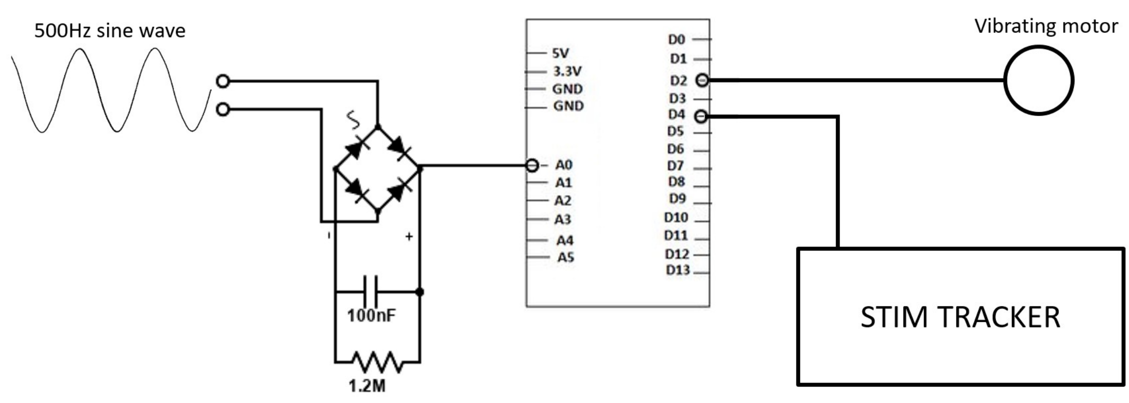

Figure 4: The Arduino circuit used to produce the tactile stimulation

2.1.4 Data Analysis

All data analysis was performed in MATLAB (R2018b; The MathWorks, Inc., Natick, MA, USA) utilizing the EEGLAB Toolbox (v2019.0) (Delorme and Makeig, 2004) and the ERPLAB plugin (v7.0.0) (Lopez -Calderon and Luck, 2014). Statistical analyses were conducted using JMP Pro statistical software package (v14.0; SAS Institute, Cary, NC, USA).

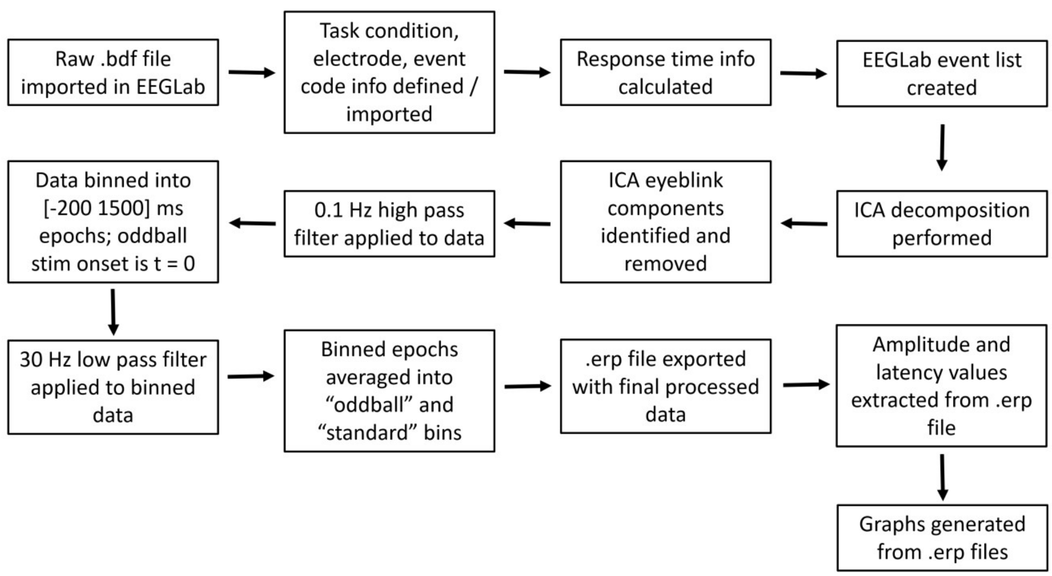

Figure 5: Flowchart of data analysis pipeline

The raw EEG data was imported into the EEGLab Toolbox and relevant information pertaining to the task conditions, electrodes, and event codes were pulled from the accompanying data files. Electrode data and event codes irrelevant to further processing were removed at this stage. Response processing for the secondary task performance was calculated at this stage as well due to the time series data being stored as integers, rather than strings in the final .erp file. The data were re-referenced to the average of the mastoid electrodes

Independent component analysis (ICA) was performed to derive a 65 channel decomposition (64 electrodes plus 2 mastoid references). In order to remove eye blink artifacts, each ICA component was compared against the vertical EOG channel and an associated correlation coefficient was calculated. Any component that had an r-value greater than 0.75 was removed. The remaining ICA channels were applied to the EEG data. At this stage, 0.1 Hz high pass filter was applied to remove frequencies associated with skin potentials.

In order to grand average across multiple trials, the raw, continuous EEG data was binned into epochs ranging from 200 ms prior to each oddball stimulus to 1500 ms post stimulus onset. One bin contained the oddball stimulus trials and another contained the standard trials. Before averaging, a 30 Hz low pass filter was applied to the binned data to remove high frequency artifacts such as muscle activity and 60 Hz line noise. The epochs were then averaged and exported as .erp files. Finally, amplitude and latency information needed for the ERP data analysis was extracted.

2.2 Tasks

Eight unique tasks were performed using one primary task and one secondary task. The primary task was either an N-back task or a compensatory tracking task. These two tasks were expressed in the visual sensory channel. The secondary task was an oddball task using either auditory or tactile stimulation. Each of the primary tasks was further categorized as either low workload or high workload. For baseline comparisons, trials where only the secondary oddball task was performed were also conducted as a control task. This brought the total task types to ten.

2.2.1 N-back

The N-back task requires participants to memorize a sequence of presented stimuli and make comparisons between the current stimulus and preceding stimuli. In this comparison, the participant must make a decision on whether the current stimulus is a target, which is defined as the current stimulus being identical to a stimulus that appeared previously in the sequence. The load factor *N* defines how far back in the sequence the current stimulus must be compared. In order to do this, the participant must store *n* stimuli in their working memory. Performance is determined by the accuracy of the participant's response in identifying targets and non-targets.

Twenty Latin letters (excluding *A, E, I, O, U,* and *Y)* were used as the N-back stimuli. 102 stimuli were presented in each task, with 34 being targets. Each stimulus was present on screen for 0.5 seconds with an inter-stimulus interval of 2.5 seconds. During each ISI, a fixation cross was presented on screen. The participant was required to provide a response with a keyboard button press to each stimuli, with one button used to denote a non-target and another to denote a target. These buttons were counterbalanced between tasks to avoid confounding effects. The low workload condition was classified as a 1-back and the high workload condition was a 3-back.

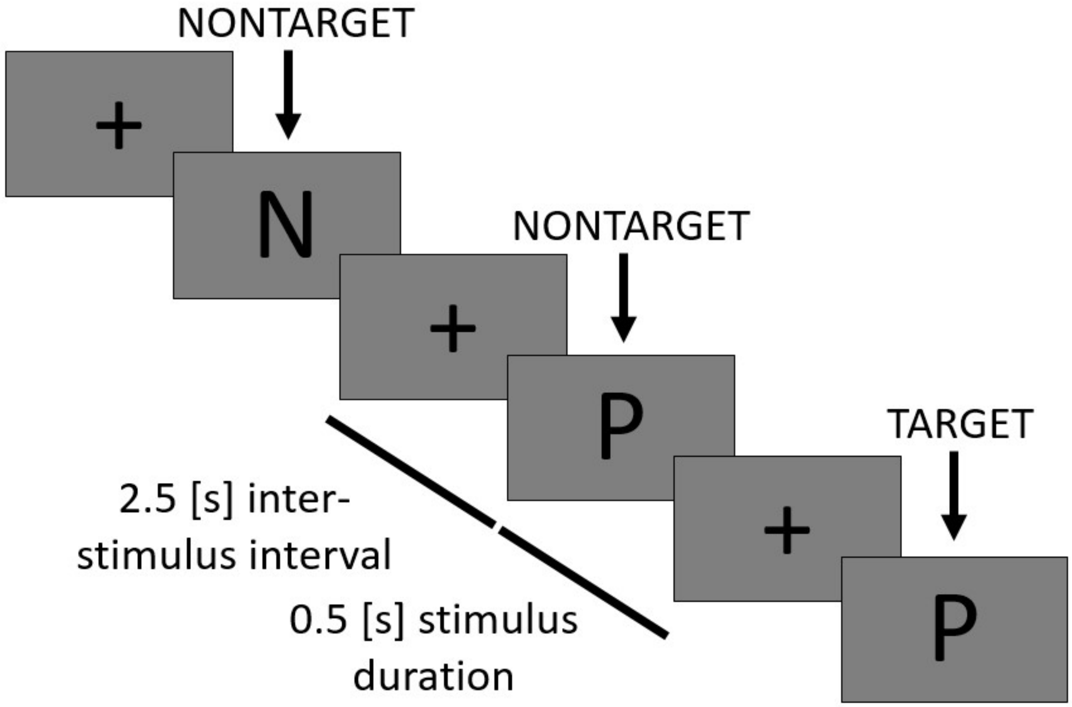

Figure 6: The N-back task presents a new alphanumeric stimulus lasting 0.5 seconds every 2.5 seconds. The participant must respond to each stimulus as either a target or non-target for the N-level given (N=1 shown).

2.2.2 Tracking

The compensatory tracking task requires a participant to follow a zero-point target around a bounded area using a cursor controlled by the game controller. When the cursor is within a sufficient distance from the target, the cursor changes color on screen from blue to green to visually indicate that the participant is on target. Performance is determined by the average distance between the two points over the course of the task.

At the beginning of each tracking task, a countdown would appear to prepare the participant to begin the task. The cursor and target begin each task at the center of the computer screen before the target travels across the screen in a continuous, random path. Each tracking task was 252 seconds long. In the low workload condition, the target moved slowly across the screen, whereas in the high workload condition, the target moved quickly and changed its direction of movement more frequently.

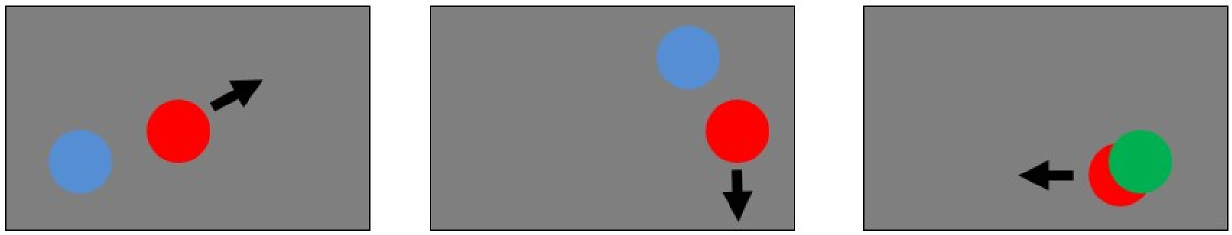

Figure 7: The tracking task has a target (red) moving randomly around the screen, and the participant must follow the target with a cursor (blue) controlled by an analog stick on a game controller. When the cursor overlaps the target, it turns green to visually indicate the desired tracking task performance to the participant.

2.2.3 Auditory Oddball

The oddball stimulus was an 800 Hz sine wave and the standard stimulus was a 500 Hz sine wave, played through the computer speakers (Logitech LS11 stereo speaker system, Logitech, Newark, CA). Each stimulus was presented for 0.15 seconds. One hundred stimuli were presented during each task, with approximately 20% of them being oddballs. To avoid confounding effects, the exact number of oddball stimuli in any single task was randomized, but the total proportion across all tasks of the same type was always 20%. The participant held a response button in their non-dominant hand which they were instructed to press whenever an oddball stimulus was presented to them.

2.2.4 Tactile Oddball

The oddball stimulus was a 500 Hz sine wave presented at 60% of maximum volume. The standard stimulus was an identical sine wave presented at 20% maximum volume. Each stimulus was presented for 0.15 seconds. The ratio of oddballs to standards was identical to the auditory oddball condition, as were the instructions to provide a response whenever the oddball was observed.

2.3 Performance Metrics

The cognitive load of each participant was evaluated using the three primary methods described above: physiological measures, specifically the P300 ERP component, task performance, and subjective evaluation.

2.3.1 ERP Data

The P300 ERP component was chosen to evaluate cognitive load based on its morphology having shown sensitivity to indexing load.

P300 peak amplitude: Peak amplitude is defined as the greatest voltage, measured in microvolts (μV) recorded during the epoched time period. As cognitive workload increases, peak amplitude is expected to decrease.

P300 peak latency: Peak latency is defined as the time at which the peak amplitude was recorded, post-stimulus onset. Peak latency is measured in milliseconds. As cognitive workload increases, peak latency is expected to remain constant.

2.3.2 Task Performance

The primary task the participant attended to occupied the visual sensory channel. The goal of these tasks was to modulate the cognitive workload the participant was experiencing. As cognitive workload increased, task performance is expected to decrease. What constitutes a decrease in performance is defined by each task.

N-back task accuracy: The participant was required to correctly identify each stimulus on screen as either a target or non-target. A response was required for each stimulus. Performance was measured by the correct identification of the current stimulus on screen, expressed as a percentage. The performance was weighted, so that the correct identification of target stimuli was a higher contributor to task performance (targets = 67% weight, non-targets = 33%). As cognitive workload increases, N-back task accuracy is expected to decrease.

Tracking task RMSE: The performance was defined as the root mean square error, which was the average relative center point distance between the target and cursor during the task. The unit is the number of pixels between the center points. As cognitive workload increases, tracking task RSME is expected to increase.

Oddball response time: This metric was defined as the amount of time that elapsed between when the oddball stimulus was presented to the participant and when the participant responded. The unit is milliseconds. As cognitive workload increases, oddball reaction time is expected to increase.

2.3.3 Subjective Evaluation

At the end of each task, participants were asked to complete the NASA Task Load Index survey to evaluate their performance. Their performance was ranked on six different scales: physical demand, mental demand, temporal demand, performance, effort, and frustration. Each scale was weighted based on which scale the participant identified as a greater contributor to their workload. The participant's responses to the survey were aggregated into a total workload score. As cognitive workload increases, total workload score is expected to increase.

2.4 Assessing Normality

Prior to statistical testing, the assumption of normality was tested to determine the quality of the ERP data. The distribution of the P300 peak amplitude data were plotted in normal quantile plots for each of the main four task categories combinations (N-back + auditory oddball, N-back + tactile oddball, tracking + auditory oddball, and tracking + tactile oddball) (figure 8).

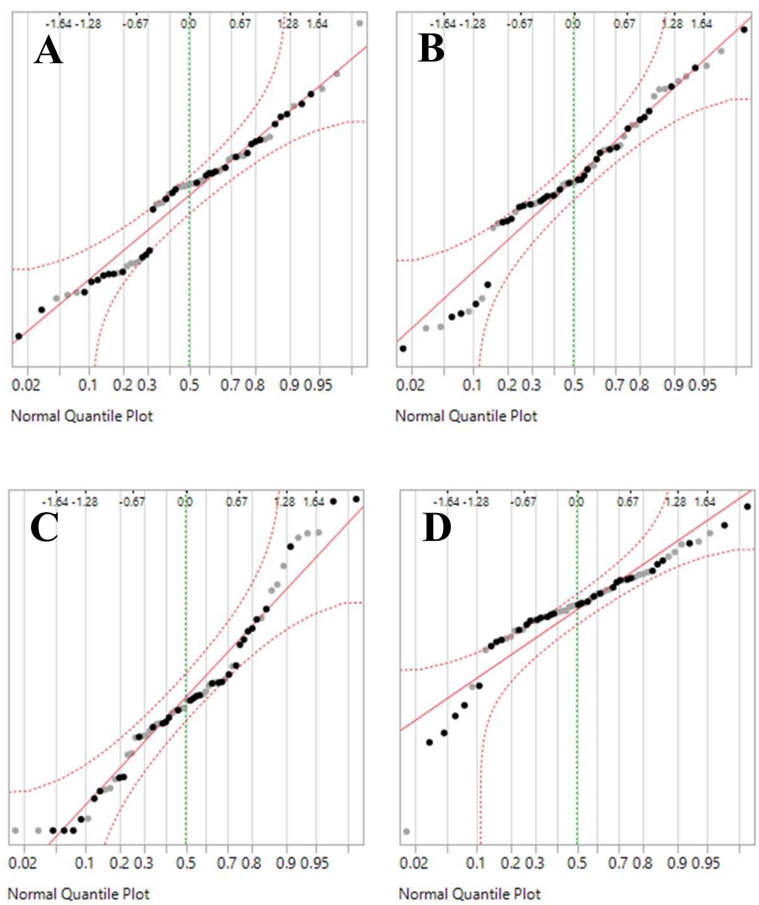

Figure 8: Normal quantiles plot for ANB (A), ATT (B), TNB (C), TTT (D)

Shapiro-Wilk Goodness-of-Fit tests were performed to determine normality and it was concluded that three of the four categories had nonparametric distributions (ANB: p = 0.0704, ATT: p = 0.0350*, TNB: p = 0.0214*, TTT: p = 0.0001).

A logarithmic transformation was used to change the data sets from nonparametric to parametric distributions. After transformation, the same Shapiro-Wilks test were ran and it was determined that all four categories no longer exhibited nonparametric distributions (ATT: p = 0.3278, TNB: p = 0.8143, TTT: p = 0.2709) (figure 9).

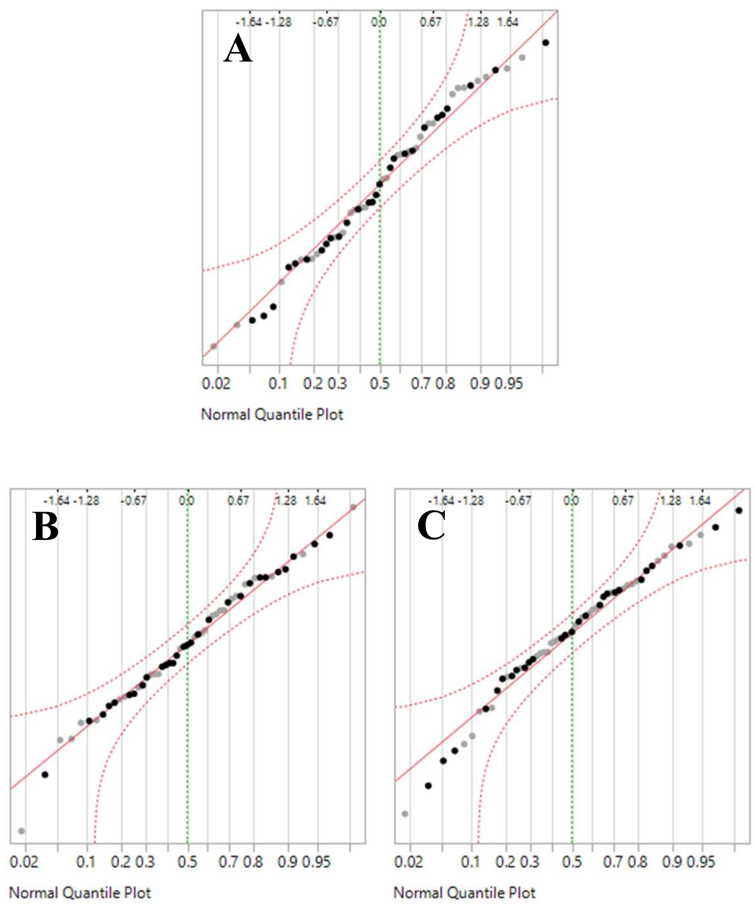

Figure 9: Normal quantile plots for log-transformed ATT (A), TNB (B), TTT (C)

III. RESULTS

All statistical results were derived from paired Student's t tests ($\alpha = 0.05$) to determine significance between different workload levels. P300 peak amplitude, N-back accuracy, and oddball response time used a left-tailed t-test. Tracking task performance and oddball response time used a right-tailed t-test. Peak latency used a two-tailed t-test. Error bars denote standard deviation. One star represents $p < 0.05$, two stars represents $p < 0.01$, and three stars represents $p < 0.001$.

3.1 Auditory Stimulus

3.1.1 ERP Data

All graphs displaying P300 components are epoched across a time 200 ms pre-stimulus onset to 1500ms post onset. Peak amplitude was at its highest in the 1-back task with 11.05 ± 15.23 uV, and then decreased in the 3-back task to 9.88 ± 15.80 uV (figure 10). No significance was found among the task difficulties ($p = 0.1980$).

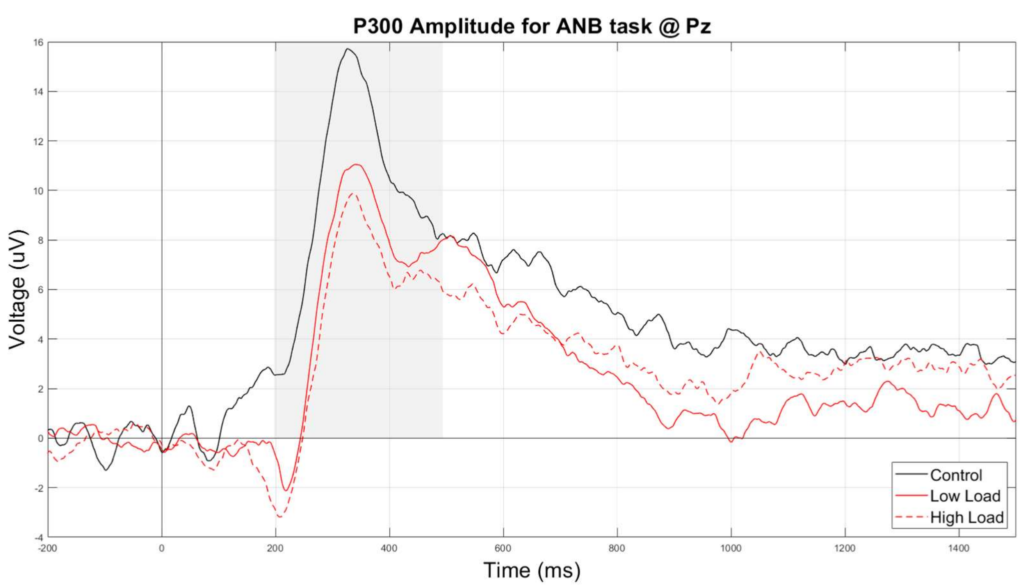

Figure 10: Grand averaged P300 component recorded at the Pz electrode during N-back + auditory oddball task

Table 1: Student's t test output for N-back + auditory oddball

t Test (High – Low)

Difference	-3.317	t Ratio	-0.85483
Std Err Dif	3.880	DF	61.91621
Confidence	0.95	P < t	0.1980

P300 amplitude was measured to be 10.19 ± 10.70 µV in the low workload tracking task, which decreased to 8.86 ± 10.62 µV in the high workload condition (figure 11). This difference was not found to be statistically significant ($p = 0.2113$).

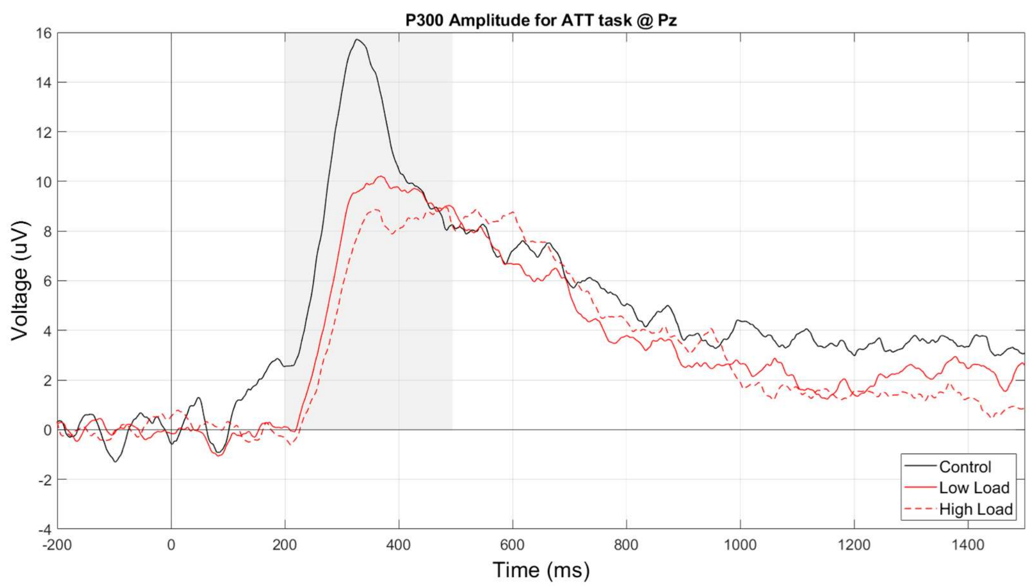

Figure 11: Grand averaged P300 component recorded at the Pz electrode during Tracking + auditory oddball task

Table 2: Student's t-test output for log-transformed tracking + auditory oddball

t Test (High – Low)

Difference	-0.04924	t Ratio	-0.8081
Std Err Dif	0.06093	DF	52.9907
Confidence	0.95	Prob < t	0.2113

Peak latency was 359.24 ± 89.53 ms in the low workload condition during the N-back task (figure 12A), and latency increased in the high workload condition to 343.89 ± 89.09 ms. No significance was found between the two task difficulties (p = 0.4943).

The tracking task showed a reversed direction of effect, decreasing from 384.06 ms in the low workload conditions to 378.83 ms (figure 12B), though this difference wasn't found to be significant (p = 0.8137).

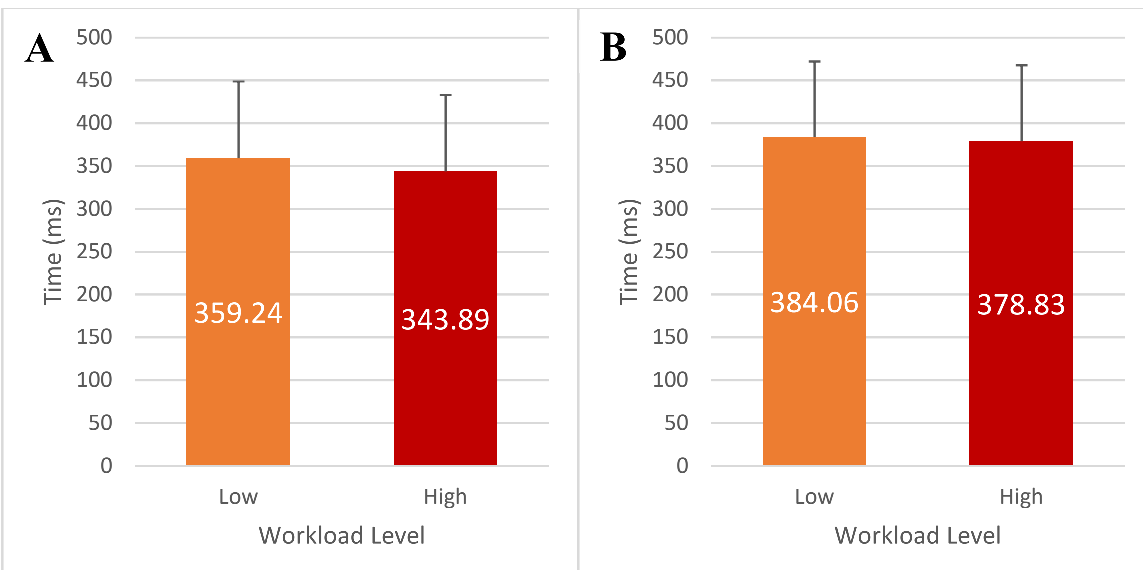

Figure 12: Post stimulus latency of P300 peak amplitude in N-back (A) and Tracking (B) tasks

Table 3: Student's t-test output for auditory N-back (left) and Tracking (right) peak latencies

t Test (High – Low)				**t Test** (High – Low)			
Difference	-15.350	t Ratio	-0.68752	Difference	-5.234	t Ratio	-0.23661
Std Err Dif	22.327	DF	61.99847	Std Err Dif	22.119	DF	61.99511
Confidence	0.95	P < |t|	0.4943	Confidence	0.95	P < |t|	0.8137

3.1.2 Task Performance

N-back accuracy in the low workload condition was 94.88 ± 3.74 percent, decreasing to 83.81 ± 16.19 percent in the high workload condition (figure 13A). This was a statistical significant decrease ($p < 0.0001$).

The average distance between target and cursor in the low workload condition was 107.22 ± 21.12 pixels, increasing to 314.54 ± 23.55 pixels in the high workload condition (figure 13B). This was a statistical significant increase ($p < 0.0001$).

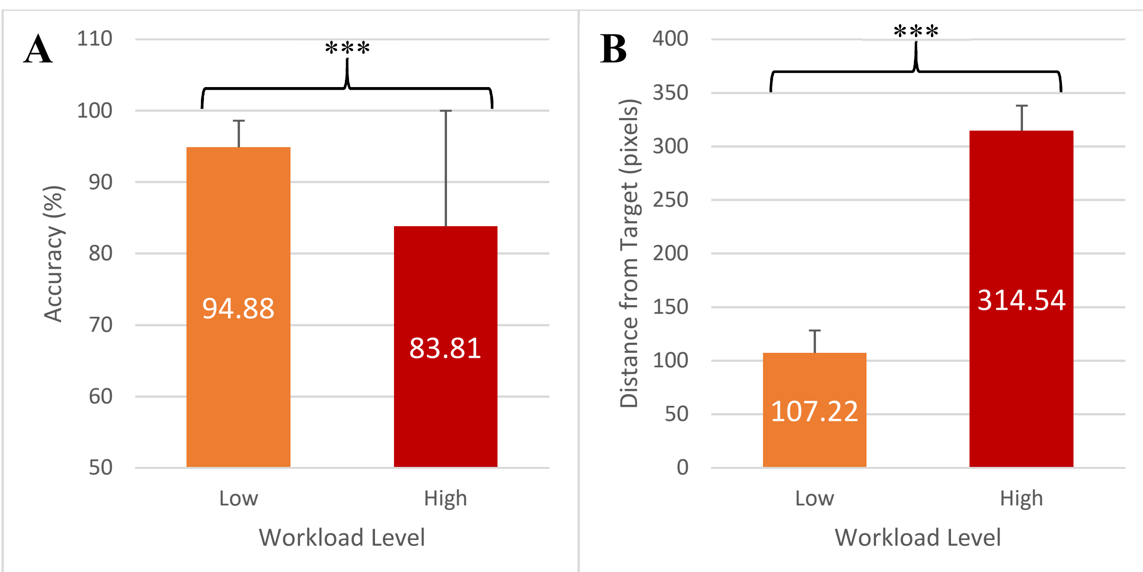

Figure 13: N-back identification accuracy (A) and the RMSE between cursor and target in the tracking task (B)

Table 4: Student's t-test output for auditory N-back (left) and Tracking (right) task performances

t Test (High – Low)

Difference	-11.063	t Ratio	-4.44716
Std Err Dif	2.488	DF	36.3792
Confidence	0.95	P < t	0.0001*

t Test (High – Low)

Difference	207.320	t Ratio	33.51195
Std Err Dif	6.186	DF	61.46812
Confidence	0.95	P > t	0.0001*

In the auditory 1-back task, participants had an average response time to target stimuli of 576.75 ± 131.15 milliseconds. When workload was increased in the 3-back (high workload) task, the time increased again to 615.44 ± 154.97 milliseconds (figure 14A). No significance was found between the low and high workload conditions ($p < 0.1104$).

When analyzing response during the tracking task, participants exhibited a response time of 521.50 ± 124.78 milliseconds in the low workload task, and 513.68 ± 144.30 milliseconds in the high load tasks (figure 14B). These results were not statistically significant ($p < 0.6305$).

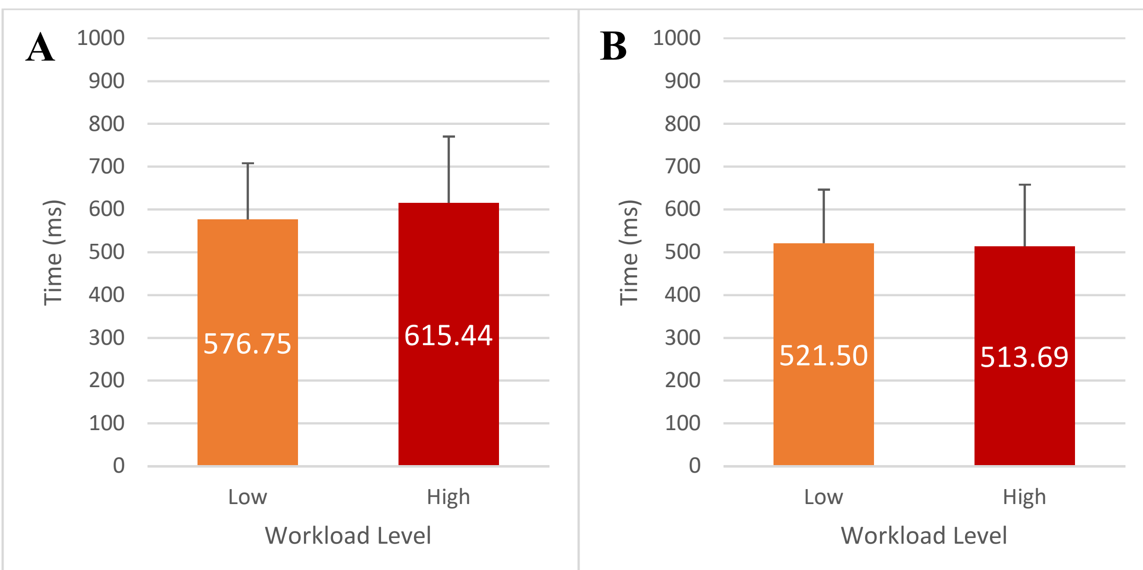

Figure 14: post stimulus oddball response times in N-back (A) and Tracking tasks (B)

Table 5: Student's t-test output for N-back (left) and Tracking (right) auditory oddball response times

t Test	(High – Low)			**t Test**	(High – Low)		
Difference	38.69	t Ratio	1.237843	Difference	-7.814	t Ratio	-0.33465
Std Err Dif	31.26	DF	58.27131	Std Err Dif	23.349	DF	61.78546
Confidence	0.95	P > t	0.1104	Confidence	0.95	P > t	0.6305

3.1.3 NASA-TLX

Evaluating participant surveys for the auditory N-back task showed a score of 38.40 ± 21.50 in the 1-back task, and the rating increased to 58.61 ± 28.84 in the 3-back task (figure 15A). The scores were found to be statistically different between the low and high workload levels ($p < 0.0012$).

In the low workload task, TLX scores for the tracking task were 32.35 ± 21.69. Once the workload was increased to high, scores also increased to 50.27 ± 29.41 (figure 15B). The task difficulties were shown to be statistically different from each other ($p < 0.0037$).

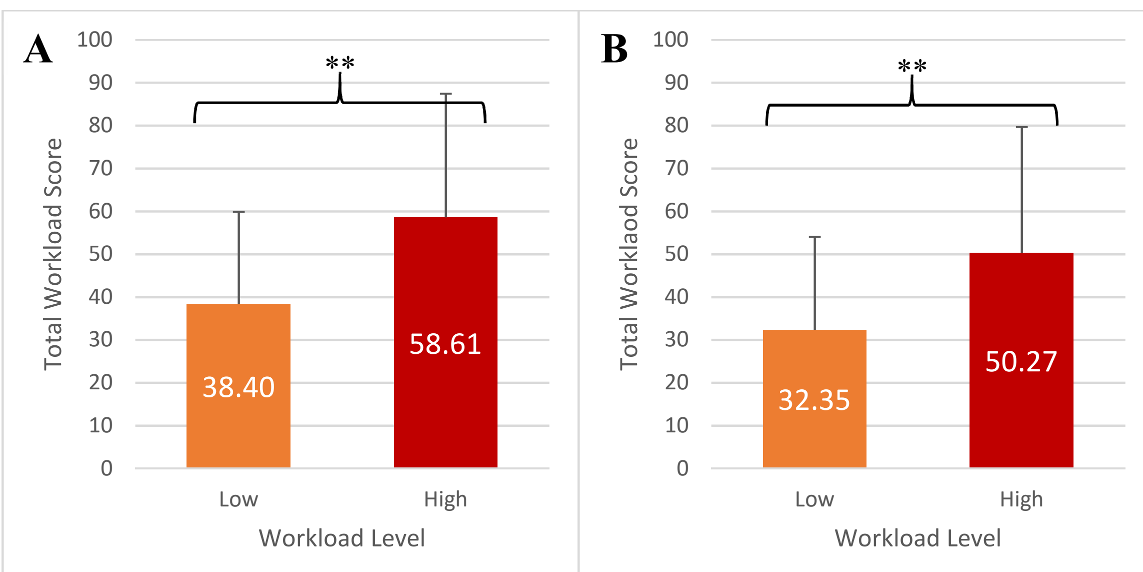

Figure 15: NASA-TLX scores recorded after the N-back + auditory oddball task (A) and Tracking + auditory oddball task (B)

Table 6: Student's t-test output for auditory N-back (left) and Tracking (right) TLX scores

t Test	(High – Low)			**t Test**	(High – Low)		
Difference	20.2188	t Ratio	3.179431	Difference	17.9167	t Ratio	2.7734
Std Err Dif	6.3592	DF	57.32872	Std Err Dif	6.4602	DF	57.02512
Confidence	0.95	P > t	0.0012	Confidence	0.95	P > t	0.0037

3.2 Tactile Stimulus

3.2.1 ERP Data

When grand-averaged across all participants, the P300 components generated by the tactile oddball tasks performed during the N-back task how a peak amplitude of 6.36 ± 8.53 μV in the low workload condition, and increased to 8.47 ± 9.42 μV when the difficulty increased (figure 16). This difference showed no significance however (p = 0.6588).

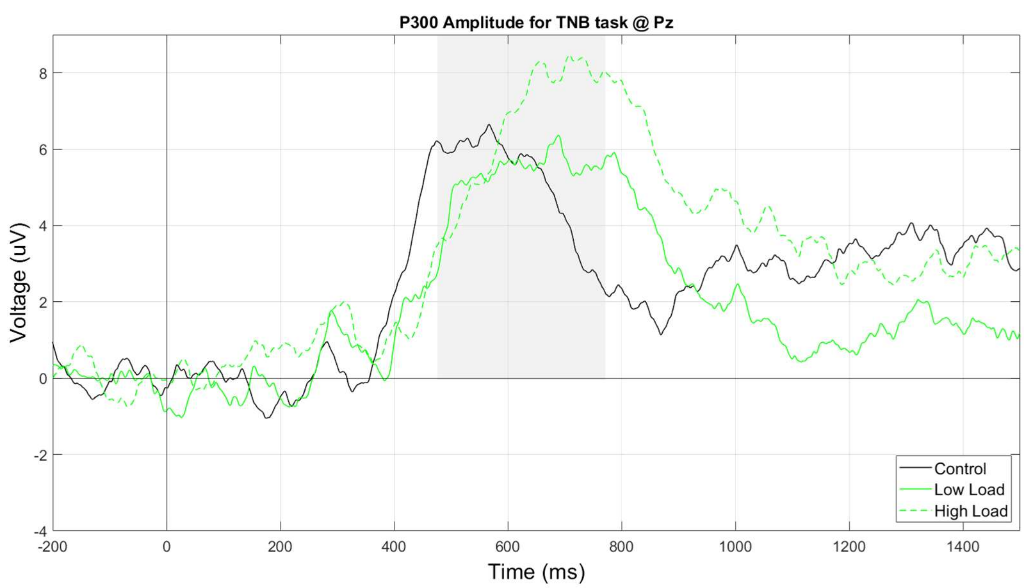

Figure 16: Grand averaged P300 component recorded at the Pz electrode during N-back + tactile oddball task

Table 7: Student's t-test output for log-transformed N-back + tactile oddball

t Test (High – Low)

Difference	0.02136	t Ratio	0.411729
Std Err Dif	0.05187	DF	49.6615
Confidence	0.95	Prob < t	0.6588

Observing the same metric in the tracking task, peak amplitude decreased from 8.81 ± 18.48 µV in the low workload group to 7.67 ± 25.91 µV in the more difficult condition (figure 17). While this downward trend was hypothesized it was not found to be significant (p = 0.4580).

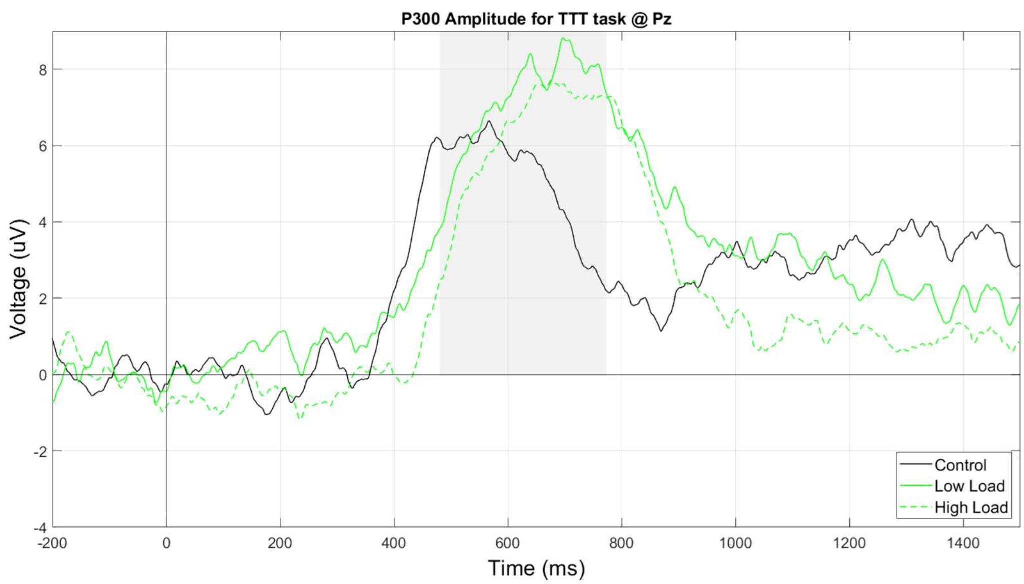

Figure 17: Grand averaged P300 component recorded at the Pz electrode during Tracking + tactile oddball task

Table 8: Student's t-test output for log-transformed tracking + tactile oddball task

t Test (High – Low)

Difference	-0.00605	t Ratio	-0.10602
Std Err Dif	0.05703	DF	53.46977
Confidence	0.95	Prob < t	0.4580

Peak latency was 656.60 ± 106.67 ms in the low workload condition during the N-back task (figure 18A), and latency increased in the high workload condition to 674.09 ± 89.92 ms No significance was found between the two task difficulties ($p = 0.4810$).

The tracking task showed a reversed direction of effect, decreasing from 384.06 ms in the low workload conditions to 378.83 ms (figure 18B), though this difference wasn't found to be significant ($p = 0.6805$).

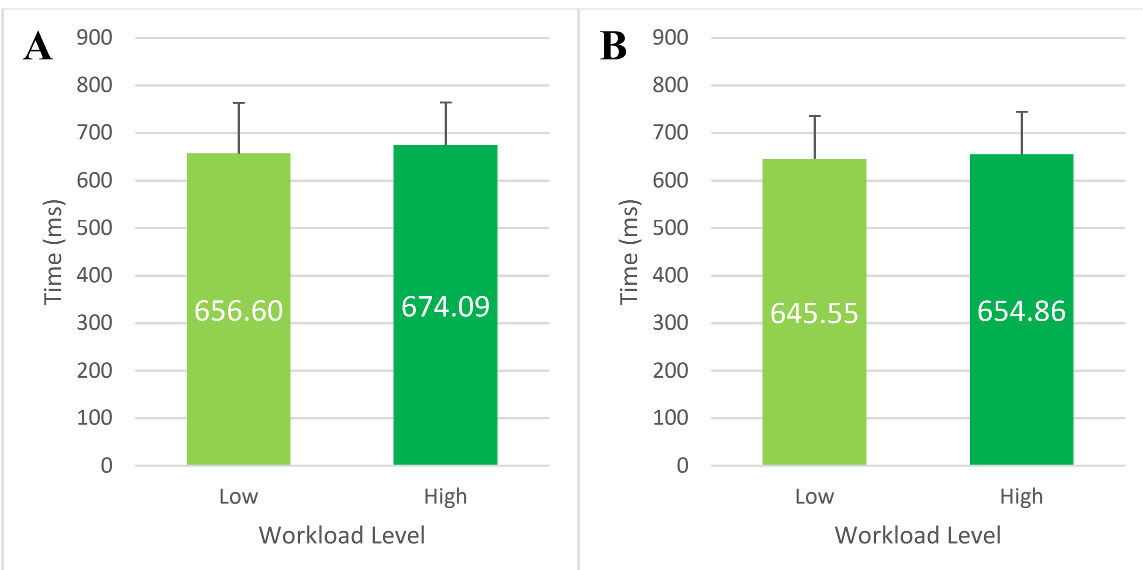

Figure 18: Post stimulus latency of P300 peak amplitude in N-back (A) and Tracking (B) tasks

Table 9: Student's t-test output for tactile peak latencies

t Test	(High – Low)			**t Test**	(High – Low)		
Difference	17.487	t Ratio	0.709048	Difference	9.308	t Ratio	-0.413767
Std Err Dif	24.662	DF	60.27467	Std Err Dif	22.495	DF	61.99754
Confidence	0.95	P > |t|	0.4810	Confidence	0.95	P > |t|	0.6805

3.2.2 Task Performance

N-back accuracy in the low workload condition was 95.44 ± 11.68 percent, decreasing to 81.07 ± 26.19 percent in the high workload condition (figure 19A). This was a statistical significant decrease ($p < 0.0001$).

The average distance between target and cursor in the low workload condition was 104.70 ± 107.33 pixels, increasing to 318.97 ± 22.98 pixels in the high workload condition (figure 19B). This was a statistical significant increase ($p < 0.0001$)

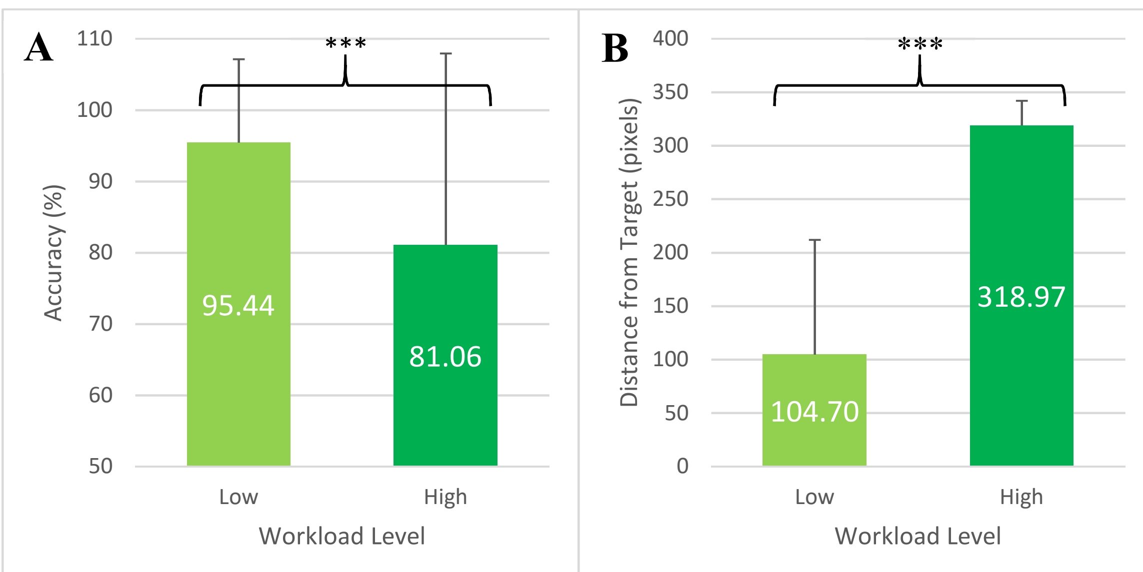

Figure 19: N-back identification accuracy (A) and the RMSE between cursor and target in the tracking task (B)

Table 10: Student t-test output for tactile N-back (left) and Tracking (right) task performances

t Test	(High – Low)			**t Test**	(High – Low)		
Difference	-14.375	t Ratio	-4.32438	Difference	214.267	t Ratio	43.48742
Std Err Dif	3.324	DF	33.30759	Std Err Dif	4.927	DF	58.7002
Confidence	0.95	P < t	0.0001	Confidence	0.95	P > t	0.0001

In the tactile 1-back task, participants had an average response time to target stimuli of 731.38 ± 138.59 milliseconds. When workload was increased in the 3-back (high workload) task, the time increased again to 780.33 ± 167.31 milliseconds (figure 20A). Significance was found between the low and high workload conditions ($p < 0.0304$).

When compared to the low workload task, (680.06 ± 129.99 ms), participants exhibited a response time of 705.47 ± 113.63 milliseconds in the high workload condition (figure 20B). No statistical difference was found ($p = 0.1596$).

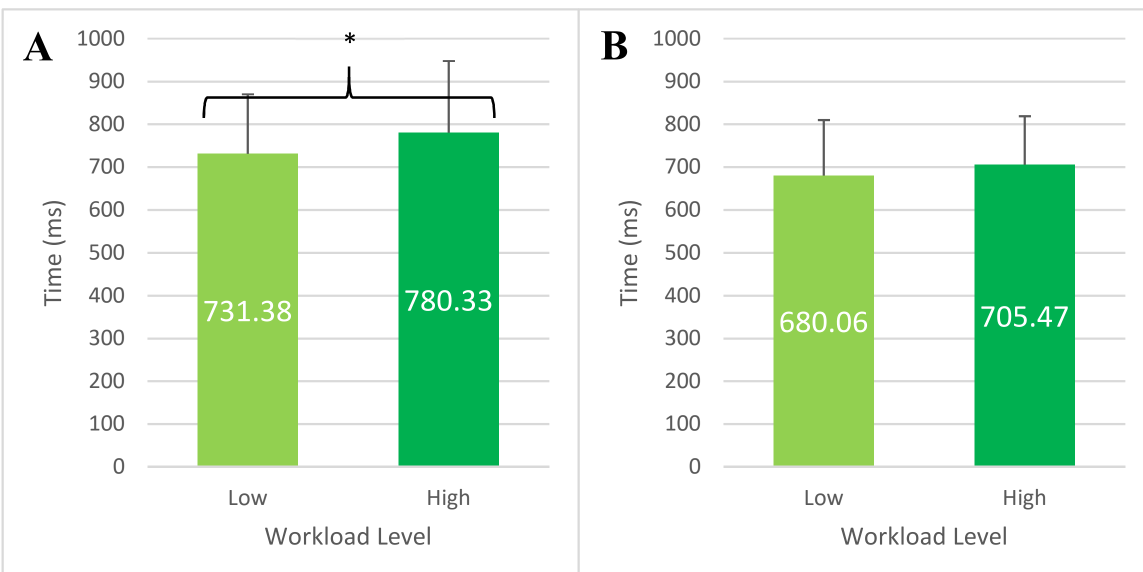

Figure 20: post stimulus oddball response times in N-back (A) and Tracking tasks (B)

Table 11: Student's t-test output for N-back (left) and Tracking (right) tactile oddball response times

t Test	(High – Low)			**t Test**	(High – Low)		
Difference	48.95	t Ratio	1.909774	Difference	25.404	t Ratio	1.004398
Std Err Dif	25.63	DF	61.6278	Std Err Dif	25.293	DF	60.48871
Confidence	0.95	P > t	0.0304	Confidence	0.95	P > t	0.1596

3.2.3 NASA-TLX

Post-task surveys generated an average score of 38.03 ± 21.69 in the tactile 1-back task. Scores increased to 60.84 ± 29.94 for the 3-back task (figure 21A). The scores were found to be statistically different between the low and high workload levels ($p < 0.0005$).

In the low workload tracking task, TLX scores were 35.90 ± 20.73. Once the workload was increased, scores also increased to 50.93 ± 30.02 (figure 21B). The task difficulties were shown to be statistically different from each other ($p < 0.0117$).

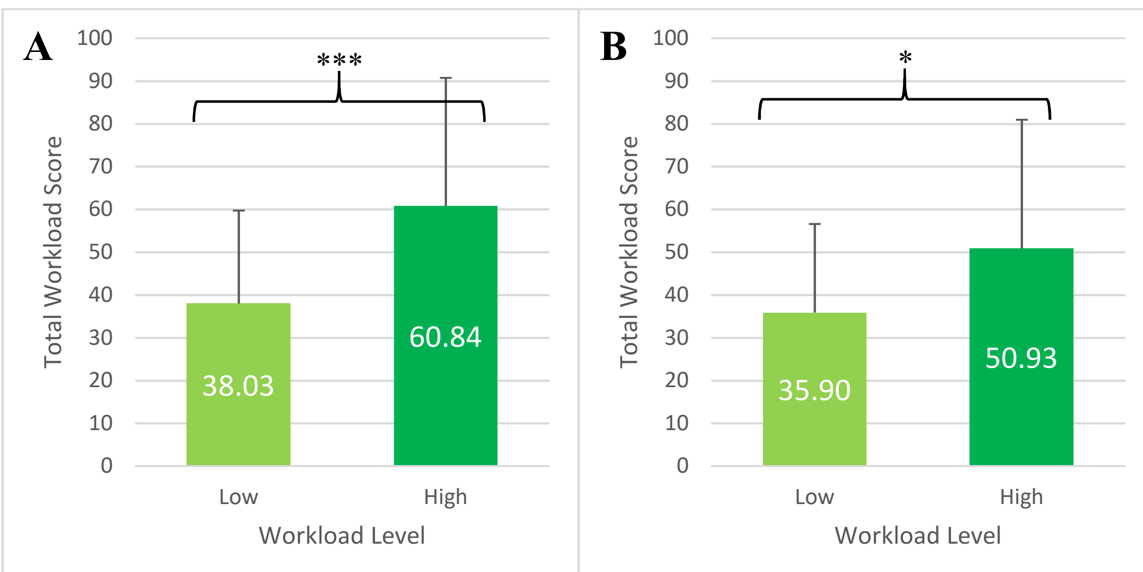

Figure 21: NASA-TLX scores recorded after the N-back + tactile oddball task (A) and Tracking + tactile oddball task (B)

Table 12: Student's t-test output for tactile N-back (left) and Tracking (right) TLX scores

t Test	(High – Low)			**t Test**	(High – Low)		
Difference	22.8125	t Ratio	3.490116	Difference	15.0312	t Ratio	2.330881
Std Err Dif	6.5363	DF	56.51575	Std Err Dif	6.4487	DF	55.08456
Confidence	0.95	P > t	0.0005	Confidence	0.95	P > t	0.0117

IV. DISCUSSION

4.1 Auditory Stimulus

4.1.1 ERP Data

The ERP data shows an attenuation in the peak amplitude of the P300 component in the N-back task. While not as substantial, a similar decrease in peak amplitude was reported in the tracking task. While this decrease was predicted in the original hypothesis, no statistical effects could be determined. Minor differences in the N-back and tracking (respectively) peak latencies were also found to be insignificant. This was expected given the understanding of P300 morphology. Peak latency is affected by the speed with which the target and non-target stimulus can be distinguished (Magliero, 1984). Since identical stimuli were presented for every task, no latency differences were expected.

It is important to note that P300 peak amplitude is sensitive to other factors besides cognitive load. For example, oddball stimulus rarity is a common variable used to modulate amplitude. This effect was mitigated in the design of the oddball task by keeping the oddball/stimulus ratio identical across all task types.

4.1.2 Task Performance

Accuracy in the N-back task dropped as the participants attempted to compensate for the increased load put on them. The tracking task showed an increased in the average tracking distance when comparing low vs high conditions. Both of these directions of effect support the hypothesized relationship with cognitive load.

The reaction times for the secondary task showed nominal changes of 6.71% and -1.50% in the N-back and tracking tasks, respectively. While these aren't significant, there is a possible explanation nested in the outlined theory governing this line of study. During task training and data collection, no attentional emphasis was placed on either the loading task or the oddball task. The oddball task is categorized as the secondary task for clarity, but that distinction from the N-back and tracking tasks is arbitrary, and MRT states that a breakdown in performance can occur in either the primary or secondary task under high load (Wickens, 1998). Maintaining performance in one task category and decrementing performance in another is supported by established theory. This was observed based on the significant differences found in N-back and tracking task performance.

4.1.3 Subjective Evaluation

The TLX scores showed a significant difference between the low and high workload conditions in both primary tasks. In the N-back task, participant scores increased between the low and high workload conditions, and an increase in the scores was found in the tracking task as well. These results support the initial hypotheses laid out for the experiment that TLX scores would increase with cognitive load.

4.2 Tactile Stimulus

4.2.1 ERP Data

The ERP data for the tactile tasks requires additional discussion. First, the N-back task does not support the expected outcome, as instead of attenuating, the P300 amplitude actually increased in the high load condition. What's more unexpected is that the tracking task data showed the reversed direction of effect, providing a decrease in P300 amplitude.

However, these opposing effect sizes cannot be considered meaningful as statistically comparison shows no difference.

The P300 peak latencies do not show any significant difference between the low and high workload conditions either. An unexpected difference was seen when comparing inter-stimulus latency values. The tactile P300s appeared to be delayed when compared to the auditory P300s. Additional analysis was performed comparing the control task data between the two stimulus methods. Peak latency for the auditory oddball was measured to be 355.67 ± 61.29 ms, but the tactile oddballs showed an average peak latency of 630.16 ± 91.96 ms, an increase of 275ms (figure 22). Literature shows that P300 ERPs tend to peak in the same time window as other sensory channels, between 300-500ms (Thurlings, Erp, Brouwer, & Werkhoven, 2013). It was not predicted that any difference between auditory and tactile stimulation would be present in this data, but when running a Student's t- tests on the auditory and tactile control data, it was found to be significantly different ($p < 0.0001$).

Figure 22: Post-stimulus latency of auditory and tactile oddball control data

Table 13: Student's t-test output for inter-stimulus latencies

t Test (High – Low)

Difference	274.490	t Ratio	14.0504
Std Err Dif	19.536	DF	54.00142
Confidence	0.95	Prob > t	0.0001

4.2.2 Task Performance

Task performance data for the tactile tasks also showed significant decreases between load conditions in the N-back task and significant increases in the tracking task. When reviewing the secondary task performance, response times increased from the low workload condition to the high workload condition by a minor amount with a significant increase detected during the N-back task. The primary task performance results were hypothesized.

4.2.3 Subjective Evaluation

When reviewing the tactile task data, a significant increase was seen in the TLX scores for both tasks. Participant responses increased from the low workload condition to the high workload condition in the N-back task. This increase was also observed in the tracking task. This increase in perceived workload was expected as the difficulty of the tasks increased.

4.4 Limitations of Study

Due to the 2020 COVID-19 health crisis, human-subjects experiments were not able to occur during the second half of this experiment's scheduled data collection. As such, only eight full data sets were included in this sample instead of the planned twenty, based on an a priori power analysis. Despite only collecting 40% of the minimum required data, significant results were still seen in several of the metrics. Significance was even shown in the most subjective evaluation: the TLX task scores. However, the data set at present is limited in its capacity to show significance in an ERP waveform. Between-subjects variability can be extremely high, which is supported by the variance reported in the ERP data, and despite seeing noticeable attenuation in the auditory tasks, no definitive conclusions could be drawn at this time.

While the auditory task results show promise, the tactile task results seem to deviate more from the predicted outcomes than had been originally envisioned. A possible explanation for this could be the manner in which the tactile stimulation was applied to the participant. The pager motor circuit was proficient in relaying the required vibrations as well as consistent time series data. However, this method differed from similar experiments also producing P300s using haptic stimulation. Further, the morphology of the component differed in terms of amplitude and latency. The use of equipment more commonly found in published literature, such as a tactor motor vest (Brouwer, van Erp, 2010; Herweg, Kubler, 2016), could produce results that resemble what was hypothesized.

V. CONCLUSIONS

5.1 Review

The risks associated with performing work under high cognitive load has been well documented, and it is an issue that has permeated the modern work culture. The consequences of such actions can range from stress, fatigue, and poor job satisfaction to more serious concerns, such as medical complications, injury, and death. Understanding how and when a person becomes overloaded in their work could elucidate solutions to overcoming this challenge.

In this study, the practical application of Multiple Resource Theory through cognitive probing was explored to determine if this method was viable for indexing cognitive load during two continuous performance tasks. To accomplish this, a secondary oddball task was deployed to target either the auditory or tactile sensory channel while the aforementioned primary task was being performed. Cognitive load was evaluated based on three measures: subjective evaluation, task performance, and physiological response.

The experimental hypothesis was supported by several of the metrics used to evaluate if differences in cognitive load could be observed. Specifically, the self-reported TLX scores increased with cognitive load, whereas primary task performance for both the N-back and tracking tasks decreased. P300 peak latency was not significantly different between the two load conditions. While not predicted in the initial hypothesis, the secondary task performance results were sensible given prior literature and theory. The P300 peak amplitude was the only metric not supported by the hypothesis or prior research.

5.2 Future Development

The research performed in this study has potential for expansion in the future. The limitations of the study prevented it from achieving a satisfactory conclusion, so returning to this line of research could prove fruitful. First, further data collection could expand the sample size to the minimum viable amount decided on. Post-hoc power analysis of the P300 peak amplitude for the four task combinations (N-back + auditory oddball, N-back + tactile oddball, tracking + auditory oddball, and tracking + tactile oddball) only revealed power estimates between 3-7%. Expanding the sample size will remove the ambiguity surrounding the current insignificant P300 amplitude results.

Further, the differences between the auditory and tactile data left one of the goals of the study unresolved, which was to generalize some relationship for cognitive load between different sensory channels. The ERP data was sufficiently different that it warranted no further analysis, but future advancements may change that. Using a standardized tactile stimulation system used in published research, such a tactor motor vest, instead of the custom circuit built for this study may produce results contained within that said research.

5.3 Acknowledgements

This project was funded by Ball Aerospace under contract to the U.S. Air Force.

REFERENCES

American Electroencephalographic Society (1994) Guideline thirteen: guidelines for standard electrode position nomenclature. American Electroencephalographic Society. J Clin Neurophysiol 11:111-113.

American Psychological Association (2018) 2018 Work and Well-being Survey

Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011). A user study of visualization effectiveness using EEG and cognitive load. In *Computer graphics forum* (Vol. 30, pp. 791–800). Wiley Online Library.

Basil, M. D. (2012). Multiple Resource Theory. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2384–2385). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-1428-6_25

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, *56*(2), 814–825.

Broadbent, D. E. (2013). *Perception and communication*. Elsevier.

Brouwer, A.-M., & Van Erp, J. B. F. (2010). A tactile P300 brain-computer interface. *Frontiers in Neuroscience*, *4*, 19.

Brouwer, A.-M., van Erp, J. B. F., Aloise, F., & Cincotti, F. (2010). Tactile, Visual, and Bimodal P300s: Could Bimodal P300s Boost BCI Performance? *SRX Neuroscience*, *2010*, 1–9. https://doi.org/10.3814/2010/967027

Burks, S. V, Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, *106*(19), 7745–7750.

Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., … Weinberger, D. R. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex*, *9*(1), 20–26.

Croft, R. J., Gonsalvez, C. J., Gabriel, C., & Barry, R. J. (2003). Target-to-target interval versus probability effects on P300 in one-and two-tone tasks. *Psychophysiology*, *40*(3), 322–328.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., & Hobart, G. (1939). Electrical reactions of the human brain to auditory stimulation during sleep. *Journal of Neurophysiology*, *2*(6), 500–514.

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*, *2*(6), 494–499.

Derrick, W. L. (1988). Dimensions of operator workload. *Human Factors*, *30*(1), 95–110.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Donchin, E. (1986). Application of brain event-related potentials to problems in engineering psychology. *Psychophysiology: Systems, Processes, and Applications*.

Fishburn, F. A., Norr, M. E., Medvedev, A. V, & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*, 76.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 301–310).

Halgren, E., Squires, N. K., Wilson, C. L., Rohrbaugh, J. W., Babb, T. L., & Crandall, P. H. (1980). Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science*, *210*(4471), 803–805.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, *7*, 83.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Herweg, A., & Kübler, A. (2016). High performance with tactile P300 BCIs. In *2016 4th International Winter Conference on Brain-Computer Interface (BCI)* (pp. 1–2). IEEE.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (1999). Analyzing and visualizing single-trial event-related potentials. In *Advances in neural information processing systems* (pp. 118–124).

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

Knight, R. T. (1984). Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *59*(1), 9–20. https://doi.org/https://doi.org/10.1016/0168-5597(84)90016-9

Knight, R. T., Grabowecky, M. F., & Scabini, D. (1995). Role of human prefrontal cortex in attention control.

Krol, Laurens R, & Zander, T. O. (2018). Cognitive and affective probing for neuroergonomics. In *Frontiers in human neuroscience conference abstract: 2nd international neuroergonomics conference* (Vol. 87).

Krol, Laurens Ruben, & Zander, T. O. (2017). Passive BCI-Based Neuroadaptive Systems, (September), 248–254. https://doi.org/10.3217/978-3-85125-533-1-46

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795.

Lindblom, J., & Thorvald, P. (2014). Towards a framework for reducing cognitive load in manufacturing personnel. *Advances in Cognitive Engineering and Neuroergonomics*, *11*, 233–244.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT press.

Marshall, S. P. (2002). The Index of Cognitive Activity: measuring cognitive workload. In *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants* (p. 7). https://doi.org/10.1109/HFPP.2002.1042860

McCarthy, G., Wood, C. C., Williamson, P. D., & Spencer, D. D. (1989). Task-dependent field potentials in human hippocampal formation. *Journal of Neuroscience*, *9*(12), 4253–4268.

McDuff, D., Gontarek, S., & Picard, R. (2014). Remote measurement of cognitive stress via heart rate variability. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2957–2960). IEEE.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*. US: American Psychological Association. https://doi.org/10.1037/h0043158

Miller, S. (2001). Workload measures. *National Advanced Driving Simulator. Iowa City, United States*.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and*

Performance, *10*(3), 358.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Puma, S., Matton, N., Paubel, P.-V., Raufaste, É., & El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology*, *123*, 111–120.

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use*. ROYAL AEROSPACE ESTABLISHMENT FARNBOROUGH (UNITED KINGDOM).

Sandblad, B., Lind, M., & Nygren, E. (1991). *Kognitiva arbetsmiljöproblem och gränssnittsdesign*. Uppsala universitet.

Segalowitz, S. J., & Barnes, K. L. (1993). The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology*, *30*(5), 451–459.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.

Stipacek, A., Grabner, R. H., Neuper, C., Fink, A., & Neubauer, A. C. (2003). Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load. *Neuroscience Letters*, *353*(3), 193–196.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Teasdale, E. L. (2006). Workplace stress. *Psychiatry*, *5*(7), 251–254.

Thurlings, M. E., Erp, J. B. F. Van, Brouwer, A., & Werkhoven, P. (2013). Controlling a Tactile ERP–BCI in a Dual Task. *IEEE Transactions on Computational Intelligence and AI in Games*, *5*(2), 129–140. https://doi.org/10.1109/TCIAIG.2013.2239294

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 3.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). An introduction to human factors engineering.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, *25*(2), 227–248.

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8), 2031–2046.

Zarjam, P., Epps, J., & Lovell, N. H. (2015). Beyond subjective self-rating: EEG signal classification of cognitive workload. *IEEE Transactions on Autonomous Mental Development*, *7*(4), 301–310.